# 传情达意：AI 合成新闻主播的身体传播刍议

赵 馨
（华东师范大学传播学院，上海 200241）

**摘 要：**智能时代下，每一次的技术突破都影响着媒介的变化和升级。人工智能技术在新闻播报领域中的运用催生了 AI 合成新闻主播。它们以全天候不间断的新闻播报为亮点走进新闻播音界，然而新闻传播所赋予的内容不仅仅是信息本身，更多的是传播过程意义下的信息，及其情感共振意义下的信息。传情达意是新闻主播基本职能概括性地表达，身体动程是达到这一传播目的的基本变化。

**关键词：**AI 合成主播；人工智能技术；新闻播报；赛博格化；新闻传播 **中图分类号：**G2 **文献标识码：**A
**文章编号：**1671-0134（2022）04-065-03 DOI：10.19483/j.cnki.11-4653/n.2022.04.018
**本文著录格式：**赵馨．传情达意：AI 合成新闻主播的身体传播刍议 [J]. 中国传媒科技，2022（04）：65-67.



新华社联合搜狗在 2018 年发布了以真人主播邱浩为原型的全球首个 AI 合成新闻主播。他依靠“搜狗分身”建模技术，将数字化信息深度整合，对合成主播进行人脸特征提取重构、唇语识别、情感迁移等多方位组合训练最终达到与真人主播无异的 AI 分身。这一所谓的技术“分身术”引起学者们多角度地观察与思考，一时之间，AI 合成新闻主播的“离身性”与真人新闻主播的“具身性”成为研究热潮，同时“传情达意”的感情交流等问题再次唤醒学界对于技术与真人的身体对比讨论，尤其是在身体互动的非语言符号方面。本文将探究 AI 合成新闻主播的身体互动传播发展历程与存在的不足，进而为真人主播提供可持续发展的方向。

## 1. AI 合成新闻主播发展概况及特点

数字媒体技术更新迭代速度不断加快，人工智能，大数据算法与媒介相互融合使得 AI 合成新闻主播应运而生。从 2018 年以邱浩为原型的 AI 合成主播到 2019 年以新华社女主播屈萌为原型打造的国内第一位 AI 合成女主播新小萌，在全国两会期间“上岗”并实现了站立式新闻播报。再到 2020 年 5 月新华社联合搜狗公司推出的以新华社记者赵琬微为原型可以高度还原真人皮肤和肌肉驱动的全球首位 3D 版 AI 合成主播新小微。AI 合成新闻主播在技术上呈现出技术迭代快、技术细致升级化的特点。三代的 AI 合成新闻主播皆以真人播音员，记者的外在特征、语音面貌为原型，进行信息采集捕捉和搭建。在新闻播报中可以做到基本的文字到声音的转换，达到与真人播音员相似的音色和规范的吐字，展现与真人主播无异的信息传达效果。

### 1.1 形象主体二重性

20 世纪 60 年代，两位美国科学家曼弗雷德．克林斯和内森．克兰在《赛博与空间》中首次提出了“赛博格”这一概念。后来这一概念延伸至定义人的身体性能经由机器拓展进而超越人体限制的新身体，简称“电子人”。国内学者彭兰在《智能时代人的数字化生存》一文中将赛博格的概念引入分析 AI 合成新闻主播：“他们都是从真人原型身上获取了面貌、声音等元件，中性地说，这些智能技术带来的是人的数字化的重组……这样一种将个体元件数字化并将各种元件重组方式，也许也成为另一种赛博格。”[1] 这样一种赛博格是一种将真人语言风格、外貌中的表情、长相甚至身材等元素符号进行数字化描写加以复制并移植到新的智能身体上。

人工智能技术当中的人脸合成技术和语音合成技术日益完善，作为赛博格化的 AI 合成新闻主播，它们在整体视觉上和听觉上都已经达到了几乎与真人主播相同的效果。尤其是在文本转化为的有声语言与动态人脸捕捉到的图像视频化配合下，显然达到了惊艳受众的逼真形象要求。在身体经验中，精神性的“自我”和物理性的“身体”都是反思后经由理性概念构造的产物，在这两者生成之前的身体经验场更为本源。在荧屏中出现的 AI 合成新闻主播显然与真人主播之间因同样的“身体经验场”产生了一种形象主体二重性的特点。

### 1.2 语音样态模糊性

从 2012 年左右，搜狗就开始针对其语音技术领域的研发和摸索，现在已逐步成为中国最大的语音识别引擎。音高、音强、音色、音长是构成语音的物理基础四要素。AI 合成新闻主播融合语音识别技术和大量的语音数据计算采集，具备各大语种、各声线、音色等语音合成的能力。在听觉效果上大约与真人主播音色一致，甚至有很强的音色辨识度效应。然而，语音的生成来源是人的发音器官，

并且它也是社会交际的工具之一。这强调了语音的来源具有具身性，即人的发音器官。强调了语音的本质属性，也是区别于其他意义上的属性，即社会性。因此，严格意义上讲，AI 合成主播所发出的声音属于一种语音识别模仿，不能称之为具有社会意义的语音。受众的听觉接收感官接收到这样的声音符号时，其技术逐渐成熟可以达到“以假乱真”的效果时，其两者语音样态就出现了一定的模糊性。

### 1.3 情感表达缺失性

当下的 AI 合成新闻主播在表达中的语言符号和非语言符号中都不断进行了优化升级，甚至在单独的非语言符号领域也探索新路径——2021 年搜狗公司发布了全球首个手语 AI 合成主播小聪，它的手势的动程基于《国家通用手语词典》，播报内容以服务于听障人士的新闻接收作为基本目标，在外部形象和手势动作方面达到逼真自然的效果。AI 合成新闻主播的打造仍然基于数字化的技术处理，这与播音学当中播音的创作方法大相径庭：所谓播音的创作方法是一种创作主体对客体进行认识、感受、理解、归纳、组织、结构的过程。在这个纬度上，情感的生发并非一种程式化的机械处理，而是涌动着的不断变化着的感情起伏。[2]AI 合成新闻主播在现阶段依然处于刚刚起步到功能优化的阶段，它的诞生和所依照的大数据恰恰反映了其情感表达上的不足，因而在这一方面考虑，人机耦合也成为目前人工合成主播应用的主流。

## 2. 中国播音学播音性质视阈下 AI 合成新闻主播的身体传播

新闻播报具有立体性、全方位的特点，它被看作是新闻宣传中的一个重要组成部分。新闻播音的性质分为两大类：创造性和多质性。其中，多质性里又包含有言语传播性、新闻性和艺术性。这些性质对研究当下 AI 合成新闻主播的身体互动具有引领作用。

### 2.1 创造性

创造性在新闻播报方面体现在创作主体发挥其主观能动性，生发出一定情感，借助情感的力量，将原先文字排列的符号系统进行改造加工，搭建起一套新的符号系统。关于创作主体，AI 合成主播和真人主播在这一方面有很大的不同。身体，是活的身体，其内在的控制、引导过程驱动着对外界的认知和互动。[3]身体是人们了解世界的媒介，人们通过身体在世界上获得各种感觉、感知，并用身体去诠释和反映独特的生命体验。动作是身体的语言，我们通过动作了解和构建世界。每天，人们会进行大量的身体动作的变化，包括肢体姿势的改变、目光的移动、肌肉的调整等，这些是人们身体动作的变化，也是日常的动作习性。AI 合成新闻主播在这里算不上是真正的具备主观能动性的创作主体。央视财经新闻主播姚雪松在向观众介绍以它为原型的 AI 合成主播姚小松时会不自觉地有着“介绍性”的身体反应，如伸手介绍、微笑等。相比之下，连线到的姚小松却略显呆板，更无任何体态语可言。

### 2.2 多质性

#### 2.2.1 语传播性质

当 AI 合成主播以与真人无异出现在荧幕上时，它们最重要也是最能表现技术的一点就在于它们的言语传播。言语传播前的传播内容的来源以及它们如何编码成声音符号的过程对研究 AI 合成新闻主播的身体互动有着内在的关联。威廉·詹姆斯在《心理学原理》中比较过两个概念“相识的知识”（Knowledge of acquaintance）和“相知的知识”（Knowledge-about）。前者依靠接触过程中的感知，直觉反应；后者则依靠概念，理性判断。这一点也就解释了为什么真人新闻主播在播报过程中为什么会自然性地伴随相应的眼神、表情和动作。AI 合成主播没有自身主体与外界的身体感知和直觉反应，它们的言语传播与真人主播的“心理—生理—心理”不同，而是“通过提取真人原型在新闻播报中的语音、唇形、表情、动作和形象等，运用智能合成技术和深度学习等联合建模数据库，将所输入的中英文文本非线性自动生成相应内容的智能音视频。”[4]目前 AI 合成新闻主播所对应的身体动作诸如眨眼、伸手、脸部肌肉张弛等都是利用多模态合成模型基于语音驱动表情序列生成对应序列的视觉效果而已。

#### 2.2.2 新闻性

国内对“新闻”这一概念的界定普遍沿用的是学者陆定一的定义：“新闻是新近发生事实的报道。”[5]在新闻播报领域，“新”要求体现出时代感、新鲜感；“事实”要求体现出新闻的真实性和分寸感；“报道”则要求主播具有一定的表态性。新闻性的表达特征也是引发身体变化的基础，比如表达激动时语速加快的同时不自觉地嘴角上扬，放大明眸；表痛心悲伤时，不自觉地皱眉等。具体再到播报不同的内容有不同的身体变化。因而，AI 合成主播在深度学习的过程中是否可以达到与真人一样精细的情感转换也有待观察。

#### 2.2.3 艺术属性

新闻播报因其具有情绪渲染下的二度创作且在听感上、视觉上兼具美的享受，因而称这一活动具有艺术属性。保罗·莱文森曾提出过媒介演进三阶段理论，前两个阶段分别为“技术作为玩具”和“技术作为现实的镜子”。他将最终的阶段比喻为“艺术”。AI 合成主播已经在技

术层面上达到玩具阶段甚至镜子阶段的水平，其仿真度越来越高。然而根据突发状况做出有声语言与自然的身体动程相匹配直击心灵的传播却难以做到。通向“传情达意”的艺术性依然道阻且长。

## 3. 传情达意：AI 合成新闻主播的未来发展

### 3.1 具身离身智能互构

媒体置于社交景观下，带来的是更强的人机互动和群际文化的交流。在强连接下，顺应媒介样态发展，注重人的生命价值，打造诗意栖息化的生活可能是未来主持传播发展的一大方向。在现实和虚拟的转换中，具身离身的智能互构体现出在互联网飞速发展时代下人类对于现实问题的反思和追求自我的渴望和尝试，构建了一个后现代的“解构重塑”的新闻传播逻辑。人机耦合凝练着对技术的创新性期待和对真实情感传达的保留。但是关于主持传播的现实和虚拟的转换搭配中，如何才是真人主持最佳的状态，如何才能达到主持传播最大化的效果，这些问题还有待进一步的探索和研究。

### 3.2 精准定位受众群体

在日趋精准导向下的媒体发展态势中，新闻主播在类别上还可以做出细化如财经新闻、国际新闻、体育新闻等，AI 合成新闻主播在受众导向方面发力不失为一条精准简明的优化路径。精准定位对不同的受众群体进行分析以及 AI 合成主播类型化打造对未来新闻传播的发展有一定积极意义。在新闻传播方面，AI 人工智能主播以庞大的信息数据作为自身的优势支撑力，在这个基础上不断将其优势放大，将信息整合、传播的单一任务逐渐朝着与真人新闻主播进行搭档、模仿等深度学习方向发展，进而发展为一种多维度服务为导向的传播方式。

### 3.3 身心交流情感沟通

“传情达意”浓缩了对新闻主播的工作内容和工作性质的要求。在人机耦合的思潮下，理性看待未来 AI 合成新闻主播的发展、洞见人机耦合的新闻传播实质对于未来人工智能走向及其应用有着重要作用。当前，人工智能在内容创作方面缺乏原创性的本质是因为人工智能是理性和“单纯”的，而人的内心世界是感性和“复杂”的。[6] 在智能时代，AI 合成主播在升级实践中集成的庞大语料可以为新闻工作者提供充足的信息保障。未来，真人新闻主播不仅是播报者的角色，而且是分析者、应用者的角色。人机协同不仅仅表现在信息处理流程层面，更多表现的又是在身心一体的深度情感沟通层面上。

## 4. 总结

就像朱广权介绍人工智能时所说：“江湖行走三尺剑六钩弓，不会武术路难行；人工智能百种智千种能，帮不到人也不成。”传播的目的是要满足人的信息需要和情感需求。本文一开始从 AI 合成新闻主播“传情达意”的硬性功能入手，顺着中国播音学的主要线索重申 AI 合成新闻主播身体动程的必要性，以及其所产生的新闻性、社会性影响。同时，立足于中国播音学中的播音性质解释了传统新闻主播的优势所在，为传统新闻主播自身提供另一条可持续发展思路。从另一角度看，结合当下的具身理论，扎根于本体实践下的中国播音学也有助于继续探索 AI 合成新闻主播的深度发展。媒

## 参考文献

[1] 彭兰．智能时代人的数字化生存——可分离的“虚拟实体”、“数字化元件”与不会消失的“具身性”[J]. 新闻记者，2019（12）：4-12.

[2] 张颂．中国播音学 [M]. 北京：北京广播学院出版社，1994：34-35.

[3] 刘海龙，束开荣．具身性与传播研究的身体观念——知觉现象学与认知科学的视角 [J]. 兰州大学学报（社会科学版），2019（2）：80-89.

[4] 於春．传播中的离身与具身：人工智能新闻主播的认知交互 [J]. 国际新界，2020（5）：35-50.

[5] 李良荣．新闻学导论 [M]. 北京：高等教育出版社，1999：14.

[6] 姜扬，宋雅琪．人工智能推动文化产业发展 [J]. 中国传媒科技，2020（1）：42-46.

**作者简介：**赵馨（1998-），女，山西太原，硕士研究生，研究方向：口语传播、视听传播、电影文化。

（责任编辑：张晓婧）